With regard to the design of major statistical surveys: are we waiting too long to evaluate substantive questionnaire content?

Esposito, James L.

Veröffentlichungsversion / Published Version
Konferenzbeitrag / conference paper

WITH REGARD TO THE DESIGN OF MAJOR STATISTICAL SURVEYS: ARE WE WAITING TOO LONG TO EVALUATE SUBSTANTIVE QUESTIONNAIRE CONTENT?[1]

JAMES L. ESPOSITO

1. Introduction

QUEST – the acronym that serves as the signature for our workshop series refers to **QU**estionnaire **E**valuation **St**andards. During each of the preceding three QUEST workshops (1997 through 2001), various attendees have written papers that address (directly or indirectly) the topic of standards for designing and evaluating survey questionnaires. One aspect of this discussion that has not been satisfactorily addressed is the point at which evaluation work actually begins. For example, does evaluation work begin formally when behavioral scientists commence cognitive testing on components of a draft questionnaire, or does/should evaluation work begin earlier with the observation and conceptualization "products" of subject-matter specialists? In the present paper, I present a framework that relates questionnaire design-and-evaluation processes to sources of measurement error and take the position that evaluation work should commence much earlier in the design (and redesign) process than has been the case historically. Though there are many excellent published works on questionnaire design and/or evaluation in the literature [Akkerboom and Dehue, 1997; Converse and Presser, 1986; DeMaio, Mathiowetz, Rothgeb, Beach and Durant, 1993; Foddy, 1993; Forsyth and Lessler, 1991; Fowler, 1995; Goldenberg, Anderson, Willimack, Freedman, Rutchik, and Moy, 2002 (for

1 The views expressed in this paper are those of the author and do not reflect the policies of the Bureau of Labor Statistics.

an establishment survey perspective); Oksenberg, Cannell and Kalton, 1991; Platek, 1985; Schwartz and Sudman, 1996; Sudman and Bradburn, 1982; Turner and Martin, 1984; and Willis, Royston and Bercini, 1991], this paper draws primarily on ideas presented by QUEST authors over the course of the past three workshops.

2. Background: Questionnaire Design-and-Evaluation Models

As noted, various members of the QUEST group have presented papers that address the topic of standards for designing and evaluating survey questionnaires, usually by presenting a descriptive model of the design-and-evaluation process. Several of those models are summarized below.

Model One. At the first QUEST workshop, Lindstrom and Akkerboom (1997) presented a "Questionnaire Testing Model" that comprised four phases:

"Phase 1, *the definition/feasibility study*, is used to construct a prototype questionnaire and data collection design usually involving a go/no go decision for further development.

Phase 2, *the qualitative content test*, is a small-scale test used mainly to produce a less error-prone questionnaire draft.

Phase 3, *the quantitative content test*, is a small-scale test used mainly to produce a less error-prone administration of the draft questionnaire, focusing on operational conditions.

Phase 4, the *quantitative pilot study*, is a final medium-scale test of the whole design (1997, pp. 10-11)"

The fourth phase is followed by the implementation phase and then the survey proper. In related papers, various authors identify and describe methods appropriate at each phase of the design-and-evaluation process. For example, during the qualitative content test, Akkerboom and Dehue (1997, Table 1a, p. 129) suggest that researchers employ the following pretesting tools: "focus groups, observation (ordinary interviews), in-depth interviews, 1-to-1 meta-interviews (cognitive interviews if using cognitive stimuli), [and] expert reappraisal." Later, during the quantitative pilot test, they suggest the use of "analysis outcomes, experiments, evaluation questions, [and] other monitoring tools (e.g., re-interviews, focus groups/debriefings)."

Model Two. At that same initial QUEST workshop, Esposito (1997) presented a five-phase model of the questionnaire design-and-evaluation process that was intended to encompass both initial design work and redesign work (see Esposito and Rothgeb, 1997).

The model comprised five phases: (1) conceptualization; (2) operationalization; (3) *pretesting* (evaluation work conducted *prior to* survey implementation); (4) survey administration; (5) *quality assessment* (evaluation work conducted *after* survey implementation). Esposito and Rothgeb (1997, pp. 543-551) also describe various evaluation methods appropriate for both pretesting and quality assessment research.

Model Three. At the 1999 QUEST workshop, Lindström presented the details of a questionnaire development-and-evaluation model that comprised seven phases:

Phase 1. Defining the contents of the survey.
Phase 2. Developing a questionnaire "at the desk".
Phase 3. Testing the questionnaire.
Phase 4. Adapting to production.
Phase 5. Monitoring the initial production.
Phase 6. Evaluation of quality and identification of sources of error.
Phase 7. Declaration of Quality

In this paper, Lindström (1999) emphasizes communication among client, methodologist and producer, describes the specific tasks associated with each phase, and identifies various methods appropriate to each phase.

Remarks and observations. Though the three models identified above differ in various respects, they (and other models of the design-and-evaluation process) seem to have accepted a dubious assumption: That *formal evaluation work* need not commence until *after* a preliminary set of questionnaire items have been developed. Formal evaluations of early developmental activity (i.e., observation and conceptualization) – such as, (1) incorporating independent/ethnographic observations of the subject-matter domain to determine how representative a sponsor's observations of that domain might be; or (2) undertaking an examination of the process by which and by whom key survey concepts have been developed – are rarely incorporated as well-developed components in questionnaire design-and-evaluation frameworks. In fact, as the following quotations suggest, there does not appear to be a great deal of enthusiasm for presurvey evaluation work of any kind.

"Selling focus groups, cognitive interviews and other pretests to survey clients is difficult. When clients make contacts, they either seem to believe that they already know how questions should be asked, or that you should be able to suggest solutions to question problems straight away. … In either case, when clients call the survey organization, they expect them to start the data collection almost instantly (Haraldsen, 1999, p. 67)."

"The competence and ambitions of our clients varies a lot. Some clients with good knowledge of measurement difficulties and with ambitions of high reliability

questionnaires will discuss in great detail their goals and their means to achieve them. They will also find resources to pay for the necessary studies and are often skilled in evaluating and using them. [New paragraph.] Often the clients do not have the time, capacity or interest to perform a questionnaire test especially adapted to their survey. When they contact the Measurement Laboratory (ML) there is a very short period available before their survey production is planned to start. Frequently these clients have designed preliminary questionnaires on their own (Henningsson, 2001, p. 73)."

"… [We] spent much time and effort this summer negotiating with questionnaire sponsors regarding questions that, as we saw in the lab, did not necessarily coincide with the actuality of respondents' lives. That is, the questions were written with an intended research agenda in mind and neglected to account for the ways in which respondents, themselves, experienced and made sense of the phenomena (Miller, 2001, p. 92)."

To better understand the importance and timing of evaluation work and its role in minimizing measurement error, we need an expanded model of the questionnaire design-and-evaluation process.

3. A Framework Relating Questionnaire Design-and-Evaluation Processes to Sources of Measurement Error

All surveys are not created equal: The framework outlined below and described very briefly is intended more for consideration in the design/redesign of interviewer-administered surveys that have recognized and ongoing societal importance (e.g., national surveys of health, employment, economic activity, social conditions, income, et cetera). A more detailed description of the framework can be found in a two recent conference papers (Esposito, 2002 and 2003).

The framework comprises two explicit dimensions, plus the implicit dimension of time/change (see Table 1):

- (1) Eight design-and-evaluation phases (for both initial design *and* redesign efforts);
- (2) Five sources of measurement error; and
- (3) The dimension of time – coupled with the inevitability of social, cultural, and technological change.

Table 1: A Framework Relating Questionnaire Design-and-Evaluation (D-and-E) Processes to Sources of Measurement Error

			INTERDEPENDENT SOURCES OF MEASUREMENT ERROR (at P7 or RP7)				
			Questionnaire D-and-E Team		Information/Data Collection Context		
		INITIAL DESIGN	*Content Specialist (1)*	*Design Specialist (2)*	*Interviewer (3)*	*Respondent (4)*	*Mode (5)*
	P1	*Observation*	C_{11}	C_{12}	C_{13}	C_{14}	
	P2	Evaluation	C_{21}	C_{22}	C_{23}	C_{24}	
Questionnaire	P3	*Conceptualization*	C_{31}	C_{32}	C_{33}	C_{34}	
Design and	P4	Evaluation	C_{41}	C_{42}	C_{43}	C_{44}	
Evaluation	P5	*Operationalization*	C_{51}	C_{52}	C_{53}	C_{54}	C_{55}
Phases	P6	Evaluation	C_{61}	C_{62}	C_{63}	C_{64}	C_{65}
	P7	***Administration***	$\mathbf{C_{71}}$	$\mathbf{C_{72}}$	$\mathbf{C_{73}}$	$\mathbf{C_{74}}$	$\mathbf{C_{75}}$
	P8	Evaluation	C_{81}	C_{82}	C_{83}	C_{84}	C_{85}
		REDESIGN					
	RP1	*Observation*	C_{R11}	C_{R12}	C_{R13}	C_{R14}	
	RP2	Evaluation	C_{R21}	C_{R22}	C_{R23}	C_{R24}	
Questionnaire	RP3	*Conceptualization*	C_{R31}	C_{R32}	C_{R33}	C_{R34}	
***Redesign* and**	RP4	Evaluation	C_{R41}	C_{R42}	C_{R43}	C_{R44}	
Evaluation	RP5	*Operationalization*	C_{R51}	C_{R52}	C_{R53}	C_{R54}	C_{R55}
Phases	RP6	Evaluation	C_{R61}	C_{R62}	C_{R63}	C_{R64}	C_{R65}
	RP7	***Administration***	$\mathbf{C_{R71}}$	$\mathbf{C_{R72}}$	$\mathbf{C_{R73}}$	$\mathbf{C_{R74}}$	$\mathbf{C_{R75}}$
	RP8	Evaluation	C_{R81}	C_{R82}	C_{R83}	C_{R84}	C_{R85}

With regard to the first dimension, four core design phases/processes are specified:

- **P1:** *Observation.* The foundation upon which survey concepts are built. Quality threats: Preconceived ideas/theories. Limited field of observation.
- **P3:** *Conceptualization.* The process of simplifying/organizing domain-relevant observations. The substantive elements upon which questionnaire items and metadata are built. Quality threats: Preconceived ideas/theories.
- **P5:** *Operationalization.* The translation of domain-relevant concepts into questionnaire items and metadata. Quality threats: Inadequate design skills.
- **P7:** *Administration.* Gathering self-report data by means of an interviewer-administered questionnaire. Quality threats: Sources of measurement error. Inadequate resources (staff and funding).

And four accompanying evaluation phases or processes:

- **P2:** Assessment of Observation Phase
- **P4:** Assessment of Conceptualization Phase
- **P6:** Assessment of Operationalization Phase
- **P8:** Assessment of Administration Phase

With regard to the second dimension, which draws largely on the work of Groves (1987, 1989), five sources of measurement error are specified (for details on the first two sources, which differ from Groves, see Esposito, 2002 or 2003):

- **S1:** Questionnaire: *Content Specialist* [subject-matter experts within a particular domain (e.g., health; labor-force dynamics; income and wealth; demographics)]
- **S2:** Questionnaire: *Design Specialist* (professionals who, in collaboration with content specialists, design questionnaires and develop ancillary metadata)
- **S3:** Interviewer
- **S4:** Respondent
- **S5:** Mode

Several additional aspects of the framework are worthy of note:

"First, it is presumed that design-and-evaluation work can and often does overlap across phases and that movement between certain phases (P1 through P6) is bidirectional and potentially iterative.

Second, the phrase "interdependent sources of measurement error" has been adopted to reflect the view that measurement error – and accuracy, too – is presumed to be the outcome of collaborative/interactive processes involving the various sources of error identified in Table 1. Within a given data-collection context, measurement error is presumed to be a byproduct of role- and task-specific activities … that manifest themselves during the survey administrative phase (P7 or RP7). Various role- and task-

specific activities that are performed inadequately at prior design-and-evaluation phases (P1 through P6) can be viewed as *precursors* to measurement error.

Third, the actual performance of role- and task-specific activities – represented as generically-labeled cell entries (e.g., C_{12}) – is presumed to vary across [questionnaire] design-and-evaluation efforts. Whether a particular cell has an entry or not would depend on whether specific cell-related activities were conducted. For example, if content specialists are not involved in pretesting work conducted during the initial questionnaire design, then cell C_{61} would be left blank. Empty cells are problematic in that they represent activity or knowledge gaps that are apt to increase the locus and magnitude of measurement error.

And lastly, as noted, social, cultural and technological change also plays a crucial role in the measurement process. Unless continuously monitored and accounted for by content and design specialists, rapid change within a given target domain can have a substantial effect on measurement error (Esposito, 2003, p. 55)."

4. Discussion

In this closing section, I will address two issues that have relevance to the question asked in the title of this paper. First: Should evaluation work begin sooner than phase six (P6), the phase during which draft questionnaires are most often pretested? I would say "yes", because *technically* well-designed questionnaire items (e.g., simple/familiar wording; good structure; acceptable working-memory demands), while necessary if high-quality survey data are to be obtained, provide no guarantee that measurement error will be minimized. At the earliest stages of the development/design process, we must seek to establish domain-relevant *grounding* for all of the substantive concepts mentioned in our draft questionnaires – and we need to evaluate that foundational work, whatever the source (see suggestions below). In framework-specific terms (Table 1), we can see that there are many threats to measurement accuracy situated upstream in the early phases of the design-and-evaluation process (e.g., at P1 and P3). Evaluation work conducted during P6 cannot be expected to identify (e.g., using standard pretesting techniques) or successfully remove/avoid (e.g., via modifications to item wording or questionnaire structure) all of these potentially damaging threats. The evaluation process must begin sooner. Postponing evaluation work until P6 could prove unwise for other more pragmatic reasons. For example, sponsors/clients, wanting "hard evidence" of problems, could choose not to implement specific research recommendations due to reservations about the use of qualitative evaluation techniques (e.g., focus groups; cognitive interviews); or they could choose not to implement some recommendations due to the sheer number or

magnitude of problems detected and/or because of insufficient time or funding (see Rothgeb, Loomis and Hess, 2001). Again, the sooner problems are identified, the better the chances that they might be considered and resolved.

Regarding the second issue alluded to above: What actions might practitioners take to assure themselves that questionnaire development/design activities have been properly grounded, empirically and conceptually? Here are some suggestions:

- Request that sponsors/clients provide documentation to support their observations and conceptualizations of the target domain (see Table 1, phases P1 and P3, respectively). In the absence of sufficient metadata – and in collaboration with content specialists, we should consider gathering empirical/behavioral data that might be used to support and/or to evaluate prior observational and conceptual work, and we should make these metadata available to sponsors/clients in a timely fashion.
- During the early stages of questionnaire design, request information on the source of draft questionnaire items and carefully examine available documentation/metadata. For items taken from preexisting questionnaires and incorporated into a new questionnaire, obtain whatever documentation/metadata might be available to assess the empirical and conceptual foundations of these items, and make note of substantial conceptual disparities among draft items and transplanted items.

Because the nature of work differs at different points in the questionnaire design-and-evaluation process (e.g., domain-relevant observations and conceptualizing in natural contexts versus more restricted behavioral observations/activities in laboratory, office or field-based contexts), the research methodologies used in the early phases of that process (e.g., P2 and P4) will tend to differ from those used later (e.g., P6 and P8; for thoughtful discussions of these methods, see DeMaio et al., 1993; Forsyth and Lessler, 1991). So, to the extent that practitioners choose to act on the suggestions offered above, we must expand our research repertoires to incorporate a variety of ethnographic, sociological and social psychological methods (e.g., Beebe, 2001; Gerber, 1999; Glaser and Strauss, 1967/1999; Hox, 1997; Webb, Campbell, Schwartz, and Sechrest, 1966). One might also expect that taking action on the suggestions offered above will have a greater likelihood of success: (1) if communications between content and design specialists begins at the early planning stages of the design-and-evaluation process, rather than later (Rothgeb, Loomis and Hess, 2001), and (2) if every participating group – sponsors/clients; content, design and production specialists; interviewers and respondents – understands the essentially collaborative nature of this process.

References

Akkerboom, H., and Dehue, F., 1997: "The Dutch Model of Data Collection Development for Official Surveys." International Journal of Public Opinion Research, 9, 126-145.

Beebe, J., 2001: Rapid Assessment Process: An Introduction. Walnut Creek, CA: AltaMira Press.

Converse, J.M., and Presser, S., 1986: Survey Questions: Handcrafting the Standardized Questionnaire. Newbury Park CA: Sage.

DeMaio, T., Mathiowetz, N., Rothgeb, J., Beach, M.E., and Durant, S., 1993: Protocol for Pretesting Demographic Surveys at the Census Bureau. Washington, DC: U.S. Bureau of the Census.

Esposito, J.L., 2003: "A Framework Relating Questionnaire Design and Evaluation Processes to Sources of Measurement Error." Proceedings of the 2003 Research Conference of the Federal Committee on Statistical Methodology, Statistical Policy Working Paper Number 37. Washington, DC: U.S. Office of Management and Budget. [Available at: http://www.fcsm.gov/reports/ (at some point in 2004).]

Esposito, J.L., 2002: "Iterative, Multiple-Method Questionnaire Evaluation Research: A Case Study." Paper presented at International Conference on Questionnaire Development, Evaluation and Testing (QDET) Methods, Charlestown, SC.

Esposito, J.L., and Rothgeb, J.M., 1997: "Evaluating Survey Data: Making the Transition from Pretesting to Quality Assessment." In L. Lyberg, P. Biemer, M. Collins, E. deLeeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.), Survey Measurement and Process Quality. New York: Wiley, 541-571.

Esposito, J.L., 1997: "The Survey Design/Redesign Process: A Confluence of Art, Science and Negotiation." Proceedings of the 1997 QUEST Workshop. Örebro, Sweden: Statistics Sweden.

Foddy, W., 1993: Constructing Questions for Interviews and Questionnaires. Cambridge, UK: Cambridge University Press.

Forsyth, B.H., and Lessler, J.T., 1991: "Cognitive Laboratory Methods: A Taxonomy." In P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds.), Measurement Errors in Surveys. New York: Wiley, 393-418.

Fowler. F.J, 1995: Improving Survey Questions: Design and Evaluation. Thousand Oaks, CA: Sage.

Gerber, E., 1999: "The View from Anthropology: Ethnography and the Cognitive Interview." In M.G. Sirken, D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur, and R. Tourangeau (eds.), Cognition and Survey Research. New York: Wiley, 217-234.

Glaser, B.G., and Strauss, A.L., 1967/1999: The Discovery of Grounded Theory. New York: Aldine de Gruyter.

Goldenberg, K.L., Anderson, A.E., Willimack, D.K., Freedman, S.R., Rutchik, R.H., Moy, L.M., 2002: "Experiences Implementing Establishment Survey Questionnaire Development and Testing at Selected U.S. Government Agencies." Paper presented at International Conference on Questionnaire Development, Evaluation and Testing (QDET) Methods, Charlestown, SC.

Groves, R.M., 1989: Survey Errors and Survey Costs. New York: Wiley.

Groves, R.M., 1987: "Research on Survey Data Quality." Public Opinion Quarterly, 51, S156-S172.

Haraldsen, G., 1999: "Identifying Clients Needs for Questionnaire Development by a Pre-Survey Questionnaire." Proceedings of the 1999 QUEST Workshop. London: Office of National Statistics.

Henningsson, B., 2001: "An Enlightened Client." Proceedings of the 2001 QUEST Workshop. Washington, DC: U.S. Census Bureau.

Hox, J.J., 1997 : "From Theoretical Concept to Survey Question." In L. Lyberg, P. Biemer, M. Collins, E. deLeeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.), Survey Measurement and Process Quality. New York: Wiley, 47-69.

Lindström, H.L., 1999: "Client – Methodologist – Producer Communication to Promote Questionnaire Development and Evaluation." Proceedings of the 1999 QUEST Workshop. London: Office of National Statistics.

Lindström, H.L. and Akkerboom, H., 1997: "Terminology and Abbreviations for the Workshop." Proceedings of the 1997 QUEST Workshop. Örebro, Sweden: Statistics Sweden.

Miller, K., 2001: "Making the Sponsor – Respondent Link in Questionnaire Design." Proceedings of the 2001 QUEST Workshop. Washington, DC: U.S. Census Bureau.

Oksenberg, L., Cannell, C., and Kalton, G., 1991: "New Strategies for Pretesting Questionnaires." Journal of Official Statistics, 7, 349-365.

Platek, R., 1985: "Some Important Issues in Questionnaire Development." Journal of Official Statistics, 1, 119-136.

Rothgeb, J.M., Loomis, L.S., and Hess, J.C., 2001: "Challenges and Strategies in Gaining Acceptance of Research Results from Cognitive Questionnaire Testing." Proceedings of the 2001 QUEST Workshop. Washington, DC: U.S. Census Bureau.

Sudman, S., and Bradburn, N.M., 1982: Asking Questions: A Practical Guide to Questionnaire Design. San Francisco: Jossey-Bass.

Schwarz, N., and Sudman, S. (eds.), 1996: Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research. San Francisco: Jossey-Bass.

Turner, C.F., and Martin, E., 1984: Surveying Subjective Phenomena, Volume 1. New York: Russell Sage Foundation.

Webb, E.J., Campbell, D.T., Schwartz, R.R., and Sechrest, L., 1966: Unobtrusive Measures: Non-reactive Research in the Social Sciences. Chicago: Rand McNally.

Willis, G.B., Royston, P., and Bercini. D., 1991: "The Use of Verbal Report Methods in the Development and Testing of Survey Questionnaires." Applied Cognitive Psychology, 5, 251-267.

Contact

James L. Esposito
Bureau of Labor Statistics
Postal Square Building
2 Massachusetts Ave., NE
Washington, DC 20212
U.S.A.
email: Esposito.Jim@bls.gov